The Effects of Iliad on Medical Student Problem Solving

Charles W. Turner, PhD (1), John Williamson, MD (3, 4), Michael J. Lincoln, MD (2, 3, 4), Peter J. Haug, MD (2, 3), James Buchanan, MD (3), Curtis Anderson, MS (2), Morgan Grant, BS (1), Robert Cundick, PhD (2), and Homer R. Warner MD, PhD (2,3)

University of Utah: Departments of Psychology (1), Medical Informatics (2), School of Medicine (3), Salt Lake City Veteran's Administration Medical Center (4)

Abstract

The present study examined the effects of the Iliad expert system on diagnostic problem solving of third-year ($\underline{n}$ = 97) medical students. Students used Iliad to work-up simulated cases to supplement the education they received in their medicine clerkship. The results of the research provided evidence that the Iliad expert system did improve student diagnostic problem solving and decision making.

Introduction

Background

The need for improving medical diagnostic problem-solving is exceedingly important and a topic of recent inquiry [1]. Likewise, the use of simulated patient problems has long been a potential resource for managing this educational need, but has been subject of study in terms of scoring systems, external validation and use by undergraduate medical students [2-4]. Students traditionally learn diagnostic problem solving by working up many patients and gradually internalizing diagnostic rules, formulae and judgment relevant to these cases.

Unfortunately, the students' experience is limited by the number, variety, and quality of patient case examples available. Because of differences between hospital populations and changes in payment for medical care, teaching hospitals now contain a patient population that is quite different from the populations most students will eventually face in practice. Therefore, this experience may not prepare students adequately to solve the types of diagnostic problems they must eventually face in their practices. By means of expert systems and simulated patients, it might be possible to provide medical students a more adequate diagnostic problem solving education. The purpose of this study is to examine the effect of a recently developed expert system called Iliad on improving diagnostic problem solving performance of third year medical students.

Description of the Iliad system

Iliad is an expert system for medicine which may be used to provide both expert diagnostic consultations and patient simulations. [5,6,7]. The consultation mode allows students to enter the findings (symptoms, physical signs, and laboratory test results) for actual patients. Iliad provides students with carefully constructed opportunities to make diagnostic decisions based upon this patient data. The system applies information in a medical "knowledge base" to the patient information, provides a differential diagnosis, and allows the students to access the logical structure of the interpretations and decisions contained in the differential diagnosis. In the consultation mode, Iliad provides students with advice about efficiency and proficiency of their work up of the patient (taking into account the information gain and cost of alternative diagnostic procedures). Finally, Iliad also links these decisions to the appropriate medical literature. In the simulation mode, Iliad uses the information in the knowledge base to realistically simulate patient cases which the students may not otherwise see in real life. Through simulations, the faculty can provide the students with carefully selected cases designed to supplement the real case examples seen in the hospital.

Testing Iliad's educational efficacy

The purpose of this research was to determine whether Iliad is able to teach medical students better diagnostic and problem solving skills. Iliad's two modes can be used to both improve the diagnostic learning on actual patient cases and provide additional, simulated patient cases. The present research evaluated the effects of adding Iliad simulation experience to standard methods of teaching junior medical students rotating on internal medicine. Iliad simulations were used both in a learning mode (to train problem solving) and in a testing mode (to assess junior problem solving abilities). The testing mode does not allow the student to use the simulator mode's learning tools. The student's problem solving skills were tested on simulated cases for which they either did or did not have prior experience.

The experiment was designed to present students with simulated patient cases and then follow up the presentation of the cases with a test case the following week. The research hypothesis was that exposure to simulated cases in the learning mode would result in enhanced problem solving performance on similar, but disguised, cases presented in the testing mode the following week. For instance, a simulation case of pulmonary embolus would be disguised by giving it a different chief complaint, patient sex, and age during subsequent test sessions. The learning and test cases were presented to the students in a counter-balanced order so that the testing sequence or week of the clerkship rotation was not confounded with the specific medical diagnosis to be evaluated.

Method

Subjects

The subjects were all third year medical students in the 1989-1990 ($\underline{n}$ = 97) class at the University of Utah; the students participated in the research during their first medicine rotation. Three additional students failed to complete their clerkship duties and did not participate in the research. The student clerkship involves four different (i.e., successive) six week rotations. One fourth of the students are assigned to the medicine clerkship during each of these six week rotations. In the other rotations, students may complete assignments in surgery, pediatrics, psychiatry, or obstetrics and gynecology. Thus, at any one time, one fourth of the students were involved in the research. The present evaluation study examined only the first of their two six-week, Internal Medicine clerkship rotations. All of these medical students have received some computer-assisted instruction during their second year of medical school. Thus, they were familiar with computer based training. For example, they all have the opportunity to use the interactive video-disc program "Slice of Life" during their second year courses [8]. The students completed their clerkships at one of three teaching hospitals affiliated with the University of Utah Medical School: the LDS Hospital, the University of Utah Medical Center, and the Salt Lake VA Medical Center

Experimental Design

The experimental design was a 2 x 2 x 2 (Simulation Training Set x Test Case Training x Weeks) mixed factorial design [10]. The first factor was a between subjects (uncorrelated) factor while the last two variables were within subjects (correlated, repeated measures) factors. The Training Set (Rare-Common) independent variable refers to the type of cases that the students were randomly assigned to receive during their simulation training. These cases either had a relatively low prevalence (Rare: Graves disease, Addison's disease, pulmonary tuberculosis, achalasia, or ulcerative colitis) or a relatively high prevalence (Common: mycoplasma pneumonia, duodenal ulcer, chronic renal failure, diverticulitis, hypothyroidism) in teaching hospital patient populations.

The Test Case Training independent variable refers to the types of test cases assigned to the students. All students completed two trained and two untrained test cases. The Weeks independent variable refers to the time in the rotation during which the cases were presented. All students received a trained and an untrained test case during weeks 2 and 3 and again during weeks 4 and 5. The actual diagnosis and the order of trained-untrained cases were presented in different, counterbalanced random orders (i.e, a Latin square design) [10]. The patient test case in the first week for all students was acute cholecystitis; this case served as a baseline assessment.

Each student was tested six times during the clerkship rotation. Three dependent variables were collected for each test case. The first measure assessed the quality of the student's sequential hypothesis generation. This measured compared the student's working hypothesis at each step in the work-up to Iliad's diagnosis at that point in the work-up. A second dependent variable measured the accuracy of the student's final diagnosis.

Procedure

All third year medical students received a 1.5 hour orientation, which occurred on the first day of the rotation, in the use of the Iliad as part of their normal clerkship orientation. In addition, a medical faculty member was assigned to each hospital to assist students with the use of Iliad on a daily basis. A faculty member also met with all of the students once a week (following medical "Grand Rounds") to answer questions about the Iliad system. During the six week rotation, all of the students were exposed to three different types of Iliad procedures. First, each student was required by the Clerkship director to enter one of their real patients each week into the Consultation mode of the Iliad knowledge base. In addition, all students were required to complete at least one simulated patient using Iliad during each week of the rotation. Finally, each student was required to complete one test case each week using the test mode of the Iliad simulator.

Iliad's decision logic. In the consultation mode, the students use the Iliad program to enter medical findings from an actual patient [5]. Iliad applies a problem-solving heuristic to the patient findings and identifies the most likely diseases (differential diagnosis) that would produce the specific pattern of findings. Iliad allows the student to examine and participate in the diagnostic process. For example, suppose that the chief complaint is chest pain. Iliad considers what type of chest pain is present (e.g., infarction pain, typical angina, pleurisy, and chest wall pain). Each type of pain can be diagnosed according to a sets of rules developed by medical experts. Iliad contains these rules in logical units called "frames." The system integrates various findings to reach a diagnosis using both Boolean and Bayesian logic [5]. Conditionally dependent findings, which usually represent common pathophysiologic processes, are combined within a cluster using Boolean logic [6,7]. The decisions derived from the cluster logic is then passed to a Bayesian frame which combines the individual findings and cluster findings using a sequential Bayesian calculation [9].

Iliad is written in C and runs on a Macintosh computer. To reduce barriers to the use of the program, Macintosh SE/30 computers have been placed on each of the medical wards where the students are assigned. These machines have two megabytes of RAM and at least a 20 megabyte hard disk. Each computer also has a printer attached which can print any part of the case or the medical logic at any stage of the work-up. Thus, the students do not need to leave the ward in order to use Iliad. The consultations, simulations, and test cases are all completed by students while they are on the wards.

Simulation procedure. In a second mode, Iliad can create simulated patients that the student may not see on the patient wards [6]. Iliad can create realistic simulated cases by randomly simulating patient findings according to the logic contained in the frames [6]. Each simulated case was reviewed by two internal medicine faculty for validity and realism.

In the simulation mode, Iliad presents the student with a chief complaint selected from among the history findings in the simulated case. Then, the student proceeds by querying Iliad about additional historical, physical exam, and laboratory findings. The computer provides the simulated patient's response to each query. Iliad tracks the student's questioning strategy at each point in the case and compares this strategy to an "ideal" strategy. The "ideal" strategy at each point in the case is uniquely derived by determining which of the remaining available findings will provide the maximum information gain relative to the cost of obtaining the finding [6]. For each question, the student must indicate which disease is the most likely to account for the present findings. At each step in the work-up, Iliad simultaneously calculates the posterior probabilities for all plausible diagnoses. Iliad compares the student's best diagnosis to its best diagnosis at each point in the work-up. Iliad provides feedback at each step in the work-up which indicates: (1) the probability of the student's best hypothesis as compared to Iliad's best hypothesis, and (2) the efficiency of the student's questioning strategy as compared to the "ideal" strategy Iliad would have pursued.

Common and Rare diagnoses. A total of ten different simulated cases were created to represent diagnostic problems that medical students are expected to learn to solve by the end of their clerkship. Five of these simulations represented cases medical students were likely to see at some time during their third year (e.g., pneumonia). The other five simulations represented uncommon cases that students were unlikely to see (e.g., Addison's disease). These uncommon cases had been defined as important clerkship objectives by the medical faculty. These simulated cases were presented during weeks 1-5 of the first clerkship rotation. The cases are presented each week (one to each student).

Trained and untrained test cases. In each subsequent week, half of the students received a test case that was similar to the simulated case presented in the previous week ("trained" case). The remainder of the students were tested on "untrained" cases. The tests administered in the trained condition resembled previous simulations encountered in the learning mode. For instance, a student experiencing the trained condition might receive a pulmonary embolus case similar to a previous week's training simulation. The age, sex, and presenting complaints of the test case are altered so that the case initially seems very different to the student. However, the test case eventually proves to have a pattern of findings and pathophysiologic conditions that is similar to the training case. The other half of the students received an untrained case during their weekly test. In the successive weeks (2-6), the conditions were reversed so that students with trained test cases received untrained test cases and vice versa. This week-by-week reversal continued until each student had completed at least two trained and two untrained test cases.

Iliad was designed to improve the students' problem solving ability. The present evaluation research examined specific changes in problem solving associated with specific cases. However, students had additional opportunities to benefit from Iliad's use. After the students had completed their required consultation, simulation, and test case, they were able to complete additional "mystery simulations". The patient diagnoses in these additional simulations were of high prevalence for students in the Common condition and of low prevalence for students in the Rare condition. Each student completed approximately four "mystery simulations" beyond those that were required. The mystery simulations were not closely related to the training or test cases.

Testing Procedure. All students were asked to complete the six computer-based patient test cases. Iliad presents the test cases in the simulator mode, but blocks access to the learning tools used for training (e.g., reviewing the frames, examining Iliad's differential diagnosis). The students were instructed to complete each case without any assistance. They were not monitored during the test sessions. Each test case required approximately 30 minutes for completion. After all students had completed the weekly test cases, each student promptly received written feedback regarding their performance on their test case. If the student had not completed the test case, they received a letter signed by the clerkship director reminding them that they were required to complete their test cases on a timely basis. The students satisfactorily completed approximately 95% of all simulations and test cases. The students were informed that the results of each exam would be known only to the individual student and to the research team. Although the students were required to use Iliad during the rotation, their actual performance on any single activity was not disclosed to the medical faculty so as to reduce student anxiety about test performance.

Dependent variables

Primary dependent variable. The primary dependent variable was a measure of the student's sequential, step-wise reasoning about the case. The steps are defined by the sequence of patient queries (e.g., history, laboratory tests) selected by the student. Students must indicate the particular primarily or alternative diagnosis to which each query relates. These diagnoses are selected from a list of differential diagnoses maintained by the student. As more information is revealed by the queries, the student must modify this differential diagnosis (e.g., promote and demote certain diagnoses, add new diagnoses).

The students were instructed to continue working-up the simulated patient until all correct diagnoses were concluded (high level of probability sufficient to begin treatment) and all competing, but incorrect, diagnoses were excluded. After each of the student's queries about possible patient findings, Iliad uses a sequential Bayesian logic to calculate the posterior probability of the student's diagnosis and all of the leading diagnosis. These posterior probabilities are based upon the findings currently known about the case by the student at that point in the work-up. As each simulated case progresses, both Iliad's and the student's differential diagnosis change. At the end of a case, a competent student should match Iliad's differential diagnosis.

The student's sequential diagnosis score is computed following each query. This score is formed by dividing the posterior probability of the student's current diagnosis by Iliad's best diagnosis at that step in the work-up. The resulting score is multiplied by 100%. If the student's working hypothesis is the same as Iliad's then the student's score is 100%. The actual score received by the students score would be lower if the probability of their diagnosis is much lower than Iliad's current best hypothesis. The average score across all queries served as the dependent variable of sequential hypothesis making (the first three queries are "free questions" which are ignored in computing this average).

Secondary dependent variables. A second dependent variable reflected the accuracy of the student's final diagnosis. In this circumstance, the student received a score of +1 for the case if the correct diagnosis was selected at the end of the work-up. Alternatively, a 0 was assigned if the student did not have the correct diagnosis.

Results

Sequential Hypothesis scores

The student's sequential hypothesis scores for each stage in the work-up were analyzed using a 2 x 2 x 2 (Simulation Training Set x Test Case Training x Weeks) mixed factorial analysis of variance. The results of the analysis revealed a significant main effect for the Simulation Training Set [$F(1,93) = 20.40$, $p < .0001$]. Student's hypothesis scores were higher for the Rare ($M = 53.42$) than the common cases ($M = 42.01$). The analysis of variance also revealed a significant main effect for Test Case Training [$F(1,93) = 3.97$, $p < .05$]. These results indicated that students performed better in their work-up of a case if they had worked-up a similar case using the Iliad simulator in the previous week ($M = 49.78$) than if they had not seen the case the previous week ($M = 45.64$). Finally, a marginally significant main effect occurred for the Weeks independent variable [$F(1,97) = 2.97$, $p < .10$]. Students performed better late in the rotation as compared to early in the rotation. Figure 1 illustrates the means for the conditions.

Final hypothesis accuracy scores

An additional analysis was performed to clarify the results of the sequential hypothesis scores (the primary dependent

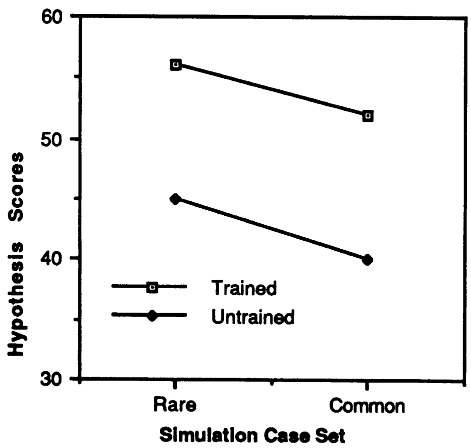

Figure 1. Effects of Simulation and Training Conditions.

variable). A 2 x 2 x 2 (Simulation Training Set x Test Case Training x Weeks) mixed factorial analysis of variance was performed using the final hypothesis accuracy scores and posterior probabilities as the dependent variables. The results for the final hypothesis accuracy score indicated that the Simulation Training Set main effect was significant, [$F(1,93) = 7.49$, $p < .007$]. Students were more likely to make a correct diagnosis for the Rare (M = 88.1%) than for the Common diagnosis (M = 77.6%). The main effect for Training approached statistical significant, [$F(1,97) = 2.85$, $p < .09$. The mean accuracy score was (M = 85.0%) for Trained cases and (M = 80.3%) for Untrained cases.

Discussion

The Iliad patient simulator provides students with a structured opportunity to practice working up patients while receiving feedback about the appropriateness of their diagnostic problem solving. The simulator can also provide experience in medical problem solving with unusual, but important, diseases. The students are unlikely to encounter these sorts of diseases among actual patients. The simulator version was designed to be especially valuable in enhancing student performance on cases with uncommon diseases. The results indicated that students achieved higher diagnostic reasoning scores on cases that they had seen previously using Iliad. This improvement occurred both for Rare and more Common diseases. The improvement in performance on the Rare cases is especially important since these cases are ones that the student is unlikely to see during their regular clerkship experience.

The experiment also demonstrated that the students' performance improves over time as the clerkship rotation proceeds. This could have occurred because the students became more familiar with the software over the course of the rotation. Alternatively, the traditional teaching methods (lectures, morning rounds, and reading) applied over the course of the rotation could have accounted for the improved performance. Further research is planned to examine these and other potential explanations.

Iliad is designed to make students aware of the relative cost and information gain for various procedures. Each query by the student during the test case can be evaluated in terms of the cost of a procedure relative to the information gain by the procedure. Although some procedures (e.g, invasive laboratory procedures) can have high information value for making a diagnosis, they may be available only at great cost. Therefore, these expensive procedures may not be optimal at an early stage in the work-up. Other procedures (including history and physical exams) may be less precise in terms of information gain, but they can be much less expensive. One possible explanation of the fact that students performed better on rare than common cases is that the students with rare cases relied more on expensive laboratory procedures (which have a higher information content) to make the diagnosis. We are currently in the process of examining the efficiency and cost effectiveness of the students' work-ups to determine if test ordering behavior differed between the Rare and Common simulated cases.

References

[1] Kassirer JP, Kopelman RI. Cognitive Errors in Diagnosis: Instantiation, Classification, and Consequences. The Am J of Med 1989; 86: 433-440.

[2] Williamson JW. Assessing Clinical Judgment; J Med Educ 1964, 40: 180-187.

[3] Melnick DE, Computer-Based Clinical Simulation: State of the Art. Eval & Health Professions, 1990; 13: 104-120.

[4] Palchik NS, Wolfe FM, et.al. Comparing Information-Gathering Strategies of Medical Students and Physicians in Diagnosing Simulated Medical Cases. Acad Med 1990, 65: 107-113.

[5] Warner HR, Haug P, Lincoln M, Warner H Jr, Sorenson D, Fan C. Iliad as an expert consultant to teach differential diagnosis. Proceedings of the 12th Symposium on Computer Applications in Medical Care, Washington, DC: IEEE Computer Society Press, 371-76, 1987.

[6] Cundick, R, Turner, C,. Lincoln, M, Buchanan, J, Anderson, C, Warner, H Jr., Bouhaddou, O. Iliad as a patient case simulator to teach medical problem solving. Proceedings of the Symposium on Computer Applications in Medical Care, 13, Washington, DC: IEEE Computer Society Press, 902-906, 1989.

[7] Turner, CW, Lincoln, MJ, Haug PJ, Warner HR , Williamson, J, Whitman, N. Clustered disease findings: aspects of expert systems. International Symposium of Medical Informatics and Education. Salamon R., Protti D., Moehr J, (eds.) University of Victoria, B.C., Canada pg. 259-263, 1989.

[8] Stensaas S, and Sorenson DK. "HyperBrain" and "Slice of Life": an interactive HyperCard and video-disk core curriculum for neuroscience. Symposium on Computer Applications in Medical Care, Washington, DC: IEEE Computer Society Press, pp 316-420, 1988.

[9] Sorenson DK, Cundick RM, Fan C, Warner HR. Passing partial information among Bayesian and Boolean frames. Proceedings of the 13th Symposium on Computer Applications in Medical Care, Washington, DC: IEEE Computer Society Press, pp 50-53, 1988.

[10] Myers JL. Fundamentals of Experimental design. Boston: Allyn & Bacon, 1979.

Acknowledgements

The research reported in this paper was supported in part by grant number 5R01-LM-046043 from the National Library of Medicine. A large number of individuals have contributed to the current research, but we are especially indebted to Dr. Richard Lee, the medical clerkship director, and to the faculty of the Department of Internal Medicine and the staff of the University of Utah medical library. Dean Sorenson, Omar Bouhaddou, Chin Li, Hong Yu, and Homer Warner, Jr. have also made valuable contributions to this research.